LWG
LOWER WILLAMETTE GROUP

PORTLAND HARBOR SUPERFUND SITE ECOLOGICAL RISK ASSESSMENT:

ESTIMATING RISKS TO BENTHIC ORGANISMS USING SEDIMENT TOXICITY TESTS

DRAFT

MAY 28, 2004

DRAFT DOCUMENT: DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

Prepared for:
The Lower Willamette Group

Prepared by:

TerraStat Consulting Group

Windward environmental LLC

Avocet Consulting

TABLE OF CONTENTS

LIST OF ACRONYMS

AET	apparent effects threshold
AFDW	ash-free dry weight
ASTM	American Society for Testing and Materials
BERA	Baseline Ecological Risk Assessment
Comp ERA tech memo	Comprehensive Synopsis of Approaches and Methods Technical Memorandum
Round 2 Comprehensive Report	Comprehensive Round 2 Site Characterization Study and Data Gap Analysis Report
COPC	chemical of potential concern
CSL	cleanup screening level
DEQ	Oregon Department of Environmental Quality
Ecology	Washington Department of Ecology
EPA	US Environmental Protection Agency
ERA	ecological risk assessment
FSP	field sampling plan
HPAH	high-molecular-weight polycyclic aromatic hydrocarbon
ISA	Initial Study Area
LPAH	low-molecular-weight polycyclic aromatic hydrocarbon
LCRMA	Lower Columbia River Management Area
LWG	Lower Willamette Group
LWR	Lower Willamette River
MDD	minimum detectable difference
NOAA	National Oceanic and Atmospheric Administration
PAH	polycyclic aromatic hydrocarbon
PCB	polychlorinated biphenyl
PEC	probable effects concentration
PEL	probable effects level

PRE	Preliminary Risk Evaluation
Programmatic Work Plan	Portland Harbor Remedial Investigation/Feasibility Study Programmatic Work Plan
QA/QC	quality assurance/quality control
QAPP	quality assurance project plan
RI/FS	Remedial Investigation/Feasibility Study
RM	river mile
SAS®	Statistical Analyses System
SEA	Striplin Environmental Associates
SPI	sediment-profile imaging
SQS	Sediment Quality Standards
SQV	sediment quality value
SVOC	semivolatile organic compound
TBT	tributyltin
TEC	threshold effects concentration
TEL	threshold effects level
TOC	total organic carbon
USACE	US Army Corps of Engineers
USGS	US Geological Survey
VOC	volatile organic compound

1.0 INTRODUCTION

The overall Ecological Risk Assessment (ERA) Approach was outlined in the *Portland Harbor Remedial Investigation/Feasibility Study Programmatic Work Plan* (Integral et al. 2004), referred to hereafter as the Programmatic Work Plan, and was approved by the US Environmental Protection Agency (EPA) in April 2004. This technical memorandum has been prepared at the request of the EPA, to describe the approach for assessing risks to benthic organisms using sediment toxicity tests. One or more predictive approaches using synoptic sediment chemistry and toxicity testing data will be developed as one line of evidence to assess risks to benthic organisms. The predictive approach that characterizes the relationship between sediment chemistry and benthic invertebrate toxicity will be developed by exploratory analyses (e.g., evaluating the predictive nature of the relationship between sediment chemistry and toxicity through comparison to existing sediment chemistry quotients or through the development of an independent predictive model). This approach will then be used to identify areas within the Initial Study Area (ISA) of the Portland Harbor Superfund Site where chemical concentrations in sediment may pose risks to benthic organisms. The results of the benthic toxicity testing will be presented in the *Results and Interpretation of the Benthic Toxicity Assessment Report.*

The assessment of risks to benthic organisms is an integral part of the ERA. If unacceptable risk is identified for a particular area or chemical of potential concern (COPC), risk reduction measures (e.g., remedial action or other methods to reduce exposure) will be evaluated. The approach presented here is the primary line of evidence in a weight-of-evidence ERA approach for assessing risks to benthic invertebrates as described in Appendix B of the Programmatic Work Plan (Integral et al. 2004). Other lines of evidence include qualitative assessment of epibenthic and infaunal invertebrate community data, sediment profile imagery (SPI) data, the tissue residue approach, and assessment of risk via the groundwater and surface water exposure pathways. This technical memorandum, however, does not address these other lines of evidence. The approach for assessing risks posed by exposure via the groundwater pathway was presented in a previous deliverable (Windward et al. 2003), and the overall approach to the assessment of risks for benthic invertebrates using all lines of evidence will be presented in the forthcoming *Portland Harbor Superfund Site Ecological Risk Assessment: Comprehensive Synopsis of Approaches and Methods Technical Memorandum* (Comp ERA tech memo).

2.0 HISTORICAL DATA

This section briefly summarizes the available Portland Harbor surface sediment chemistry and sediment toxicity test data. A more detailed description of historical data is presented in the Programmatic Work Plan (Integral et al. 2004).

2.1 AVAILABLE HISTORICAL SEDIMENT CHEMISTRY DATA

Numerous sediment investigations have been conducted within the ISA since 1990 and are summarized in the Programmatic Work Plan (Integral et al. 2004). The historical sediment data were evaluated for data quality and placed into one of two categories: Category 1 data were found to have acceptable data quality, whereas Category 2 data have an unknown, incomplete, or unsatisfactory data quality status. Detailed descriptions of the evaluation criteria and process are presented in Appendix F of the Programmatic Work Plan (Integral et al. 2004).

Sediment sampling density is the highest between river mile (RM) 4 and RM 9, where a number of facilities are undergoing remedial or other types of investigations. In addition, recent sediment chemistry analyses have also been completed for sediments just off the Gunderson Rail Car and Marine Barge Assembly Area near RM 9 (Kleinfelder 2004) and throughout the Lower Willamette River (LWR) by the US Army Corps of Engineers (USACE) as part of their dredged material characterization (Hart Crowser 2002). Figure 2-1 presents a summary of sediment data for metals, polychlorinated biphenyls (PCBs), pesticides, and total polycyclic aromatic hydrocarbons (PAHs) from the ISA based on probable effects concentrations (PECs).[1] The higher concentrations of these chemicals were measured in Terminal 4 between RM 4 and RM 5, along the western shoreline between RM 6 and RM 7, along the eastern shoreline near RM 7, and in Swan Island Lagoon (RM 8). Other chemical groups that have been analyzed (although less frequently) in ISA sediments include dioxins and furans, butyltins, petroleum hydrocarbons, semivolatile organic compounds (SVOCs), and volatile organic compounds (VOCs). In addition, data on many atypical analytes are also available.

As part of Round 1 sampling for the Portland Harbor RI/FS, surface sediment samples from 58 locations were collected and analyzed for an expanded suite of chemicals including metals, SVOCs, PCB Aroclors, and pesticides/herbicides. A subset of these

[1] PECs are consensus-based probable effects concentrations. They were developed by taking the geometric mean of existing sediment quality values (SQVs) above which adverse effects are expected (MacDonald 2000). PECs have been developed for 8 metals, 9 individual PAHs, total PAHs, total PCBs, and 9 pesticides, including total DDTs. PEC quotients were calculated for total DDTs, total PAHs, total PCBs, and each of the 8 metals by dividing the measured concentration of each by the chemical's respective PEC. A total PEC quotient was calculated as the sum of the individual quotients. Figure 2-1 presents the location for all 447 historical sediment chemistry sampling stations at which a chemical with a PEC was detected. Stations in the top 50th percentile and top 35th percentile of the total PEC quotient distribution are also indicated. Of the 447 stations included in Figure 2-1, at least one PEC was exceeded at 98 of them.

samples was also analyzed for VOCs, tributyltin (TBT), and dioxin/furan and PCB congeners.

2.2 AVAILABLE HISTORICAL SEDIMENT TOXICITY DATA

Fifteen laboratory toxicity studies using benthic invertebrates have been conducted with sediment collected within the ISA since 1990. These data were generally collected for dredged material characterizations, remedial investigations, and site investigations conducted at Portland Harbor facilities. Two tests have been commonly used to assess the suitability of dredged material for unconfined aquatic disposal: the amphipod (*Hyalella azteca*) 10-day survival test, and the midge larvae (*Chironomus tentans*) 10-day survival and growth test. In studies completed prior to the establishment of the Lower Columbia River Management Area (LCRMA) 1998 guidelines (LCRMA 1998), toxicity tests were performed using *H. azteca* and the cladoceran, *Daphnia magna*. A few studies used the Microtox® bioluminescence test.

The quality of the existing toxicity data was evaluated using a process similar to that used for the sediment chemistry data and presented in the Programmatic Work Plan (Integral et al. 2004). Based on this evaluation, twelve of the toxicity studies were designated as Category 1. In addition to these data, recent sediment toxicity studies have also been completed with sediments just off the Gunderson Rail Car and Marine Barge Assembly Area near RM 9 (Kleinfelder 2004) and throughout the LWR by the USACE as part of their dredge material management characterization (Hart Crowser 2002). In both cases, the sediment toxicity was evaluated using *H. azteca* and *C. tentans*. Figure 2-2 shows the locations where sediments have been collected from the ISA for toxicity testing using *H. azteca* and *C. tentans* from the ISA. The sediment toxicity test sampling occurred primarily in three areas along the ISA: in Slip 3, at McCormick and Baxter at RM 7, and in the vicinity of the Swan Island Lagoon at RM 8. Sediment toxicity was demonstrated at some locations within each of these three areas.

2.3 INTERPRETATION OF HISTORICAL SEDIMENT CHEMISTRY AND TOXICITY DATA

Metals and PAHs were the most frequently detected analyte groups in LWR sediments and were distributed throughout the ISA. Chlorinated pesticides were infrequently detected (detection frequency <10%), with the exception of DDT and its metabolites (i.e., 4,4'-DDE and 4,4'-DDD), 2,4-D and 2,4-DB. The detection frequencies of individual PCB Aroclors were all <30%. However, many of the historical chlorinated pesticide and PCB data had elevated detection limits and were not useful in the interpretation of sediment chemistry. PCBs and pesticides were generally analyzed only in discrete areas. Surface sampling of pesticides and PCBs occurred primarily at Terminal 4 slips, between RM 7 and RM 8, and at the Portland Shipyard. This limited sampling, along with the elevated detection limits, results in an incomplete and patchy picture of the distribution of pesticides and PCBs in the LWR. Historical data for dioxins/furans and VOCs are very limited throughout the ISA. A more representative data set will be available following the Round 2 sampling efforts of the Portland Harbor Superfund Site RI/FS (Integral and Windward 2004a).

Interpretation of toxicity test results in past studies was made by comparing amphipod and midge test results to LCRMA guidelines (LCRMA 1998). In these studies, samples tested using the *C. tentans* test failed either the growth or the mortality criterion at 5 locations in Slip 3, 1 location near RM 5, 18 locations near McCormick and Baxter (RM 7), 5 locations in the vicinity of Swan Island Lagoon, and 2 locations near RM 8. Samples tested using the *H. azteca* test failed the mortality criterion at 3 locations in Slip 3, 11 locations near McCormick and Baxter (RM 7), 15 locations in the vicinity of Swan Island Lagoon, and 1 location near RM 8. Therefore, out of a total of 128 toxicity tests for each species, there were failures for *C. tentans* and *H. azteca* at 31 and 30 stations, respectively.

2.4 USE OF HISTORICAL DATA

The historical and Round 1 data sets were used to guide the placement of Round 2a sediment and toxicity testing sampling stations throughout the ISA. The Round 2 Field Sampling Plan (FSP) (Integral and Windward 2004a) provides a complete description of how the Round 2a sediment sampling stations were selected. The historical toxicity and sediment chemistry data may be incorporated into the predictive approach, assuming the data meet data usability guidelines. One of the toxicity tests being proposed in this technical memorandum, the 10-day *C. tentans* survival and growth test, has been used historically to evaluate toxicity in Portland Harbor sediment. These data may be included when developing the predictive approach. The sediment chemistry data may be used once the approach is complete to identify areas that might be toxic based on sediment chemistry alone.

3.0 PROCESS FOR SELECTING SEDIMENT SAMPLING STATIONS

This section briefly describes the process used to select sediment sampling stations for chemical analyses and toxicity testing. A more detailed description of the entire process is included in the Round 2 FSP (Integral and Windward 2004a).

3.1 OVERVIEW OF APPROACH TO SAMPLING AND TESTING

Whole sediment chemical analyses and toxicity testing will be conducted on sediment collected from the selected locations discussed in Section 3.2. Samples will be distributed throughout the ISA to characterize the nature and extent of contamination by targeting both known major contaminant source areas and poorly characterized areas in the river. The sediment chemistry and toxicity testing data will be used to determine if a predictive relationship can be developed between sediment chemical concentrations and sediment toxicity. Such a relationship could then be used in the benthic risk assessment to predict sediment toxicity to benthic invertebrates in sediments where no toxicity tests have been conducted but sediment chemistry data are available. Details regarding how the predictive relationship will be established, and regarding the alternative approaches for areas where no relationship can be developed, are presented in Sections 7 to 9.

3.2 SAMPLING LOCATION IDENTIFICATION

In their November 11, 2003, comments on the Draft Round 2 FSP (Integral and Windward 2004a), EPA outlined the approach used for locating sediment chemistry and toxicity samples throughout the LWR. The sample locations were assigned using two approaches: grid and targeted sampling. In order to evaluate the poorly characterized areas of the river, a grid sampling plan was used to distribute sediment sample locations throughout the ISA. Using the grid sampling plan, sample locations were placed at regular intervals in the nearshore zone (<20-ft depth), at a 30-ft depth, and mid-channel throughout the ISA. To evaluate known sources, a targeted sampling plan was used to characterize major source areas by placing multiple sampling locations adjacent to known contaminant sources or other contaminant hotspots. The selected sample locations that overlapped with historical sediment sample locations were eliminated if the historical data were determined to adequately represent current conditions.

Using this approach, 223 locations were identified for sediment chemical analyses and toxicity tests. In addition to these locations, surface sediment samples will be collected at 302 additional locations for surface sediment chemical analyses alone, for a total of 525 Round 2 surface sediment sampling locations. The proposed number of paired sediment chemistry and toxicity test samples provides a large enough sample size to determine if a predictive relationship can be developed and provides the foundation for a traditional approach (e.g., additional toxicity testing) if no relationship can be developed.

A complete description of the process used to select all sediment chemistry and toxicity test locations as well as a list and map of the proposed sampling locations are presented in the Round 2 FSP (Integral and Windward 2004a).

3.3 ANALYTE LIST FOR SEDIMENT ANALYSIS

Based on assessments of the historical sediment chemistry data and consultations with EPA, 213 analytes will be included in the sediment chemistry analyses, including PAHs, PCBs, organochlorine pesticides, metals, VOCs, phenols, phthalate esters, dioxins, organonitrogen compounds, and others. The complete list is presented in Table A6-2 in the Round 2 FSP (Integral and Windward 2004a). Conventional parameters such as total organic carbon (TOC) and grain size will also be measured in each sample.

4.0 SEDIMENT TOXICITY TESTS AND TEST ENDPOINTS

This section presents the proposed sediment toxicity tests and the test endpoints that will be used for both developing the predictive approach and as measurement endpoints for the risk assessment. This section also summarizes test methods and describes the process and criteria for categorizing test results.

4.1 SELECTION OF TOXICITY TESTS AND TEST ENDPOINTS

Two toxicity tests will be conducted on the sediment samples: the chronic 28-day freshwater amphipod (*Hyalella azteca*) test, and the acute 10-day freshwater midge (*Chironomus tentans*) test. Studies of the epibenthic and infaunal communities in the LWR revealed that chironomid larvae (midges) and amphipods are major components of the benthic invertebrate community in Portland Harbor (Integral et al. 2004). *H. azteca* and *C. tentans* have been shown to be sensitive to a wide range of sediment contaminants (ASTM 2001; EPA 2000). Hence, the selected sediment toxicity tests will provide relevant site-specific information on exposure of benthic invertebrates to sediment-associated chemicals in the LWR. Both tests have an acute (survival) endpoint and a chronic (growth) endpoint that will be used to characterize the sediment toxicity in the LWR.

4.2 OVERVIEW OF TEST METHODS

4.2.1 *Hyalella* Test

The 28-day *H. azteca* sediment toxicity test will be conducted according to American Society for Testing and Materials (ASTM) Method E 1706-00 (ASTM 2003) and EPA Method 100.4 (EPA 2000). In the 28-day *H. azteca* toxicity test, amphipods are exposed to test and negative control sediments. The test is conducted with eight replicates per treatment, each containing 100 mL of sediment and 175 mL of overlying water. The test is initiated by adding ten 7- to 8-day-old amphipods to each replicate. The test chamber position is randomized, as is the distribution of amphipods to the test chambers. The test is performed at 23 ± 1°C with a photoperiod of 16L:8D. The overlying water is renewed twice daily, and the amphipods are fed once daily. At day 28, the test is terminated and the numbers of surviving amphipods in each replicate are counted and recorded. The surviving amphipods from each replicate are dried at 60 to 90°C to constant weight and weighed to the nearest 0.01 mg. The total weight of the dried amphipods from each replicate is divided by the number of amphipods weighed to obtain an average dry weight per surviving amphipod per replicate. The test is deemed acceptable if mean survival in the negative control is ≥80%.

4.2.2 *Chironomus* Test

The 10-day *C. tentans* sediment toxicity test will be conducted according to ASTM Method E 1706-00 (ASTM 2003) and EPA Method 100.2 (EPA 2000). In the 10-day

C. tentans toxicity test, larvae (midges) are exposed to test and negative control sediments. The test is conducted with eight replicates per treatment, each containing 100 mL of sediment and 175 mL of overlying water. The test is initiated by adding ten second and third instar larvae to each replicate. The test chamber position is randomized, as is the distribution of larvae to the test chambers. The test is performed at 23 ± 1°C with a photoperiod of 16L:8D. The overlying water is renewed twice daily, and the larvae are fed once daily. At day 10, the test is terminated and the numbers of surviving organisms (i.e., larvae and pupae) in each replicate are counted and recorded. The surviving larvae from each replicate (pupae are not included in the growth determination) are dried at 60-90°C to constant weight and weighed to the nearest 0.01 mg. The total weight of the dried larvae from each replicate is divided by the number of larvae weighed to obtain an average dry weight per surviving larva per replicate. The dried larvae are then ashed at 550°C for two hours. The ashed larvae are reweighed, and the tissue mass of the larvae is calculated as the difference between the weight of the dried larvae and the weight of the ashed larvae. Pupae or adult organisms are not included in the replicate to estimate ash-free dry weight (AFDW). The growth endpoint is based on the AFDW measurements. The test is deemed acceptable if mean survival in the negative control is ≥ 70% and the mean weight of surviving negative control organisms is ≥ 0.48 mg AFDW.

4.3 TOXICITY TEST RESULTS

4.3.1 Comparison to Controls

Test responses will be compared to the responses observed in the negative controls. Previous investigations with freshwater toxicity tests indicate that the negative control can be used in place of a reference sediment in interpreting toxicity responses in test sediments (ASTM 2001; Ingersoll et al. 2002). The Washington Department of Ecology (Ecology) recently evaluated the reliability of freshwater toxicity test comparisons using the negative controls against comparisons using reference stations. In general, Ecology recommends comparison to negative controls at this time because of the greater reliability[2] observed using this approach and because the results are also more conservative (Ecology 2002). Reference sediments are typically used for toxicity test comparisons to provide a measure of test response associated with non-chemical attributes of the sediment, such as sediment grain-size distribution or TOC. Hence, comparison to reference should allow field effects to be controlled for and improve the reliability of comparisons. However, based on the evaluation by Ecology (2002), this does not appear to be the case in practice for freshwater sediment. Reasons for this may include the wider variety of freshwater environments as compared to marine environments, the variability of freshwater reference stations, the current lack of identified and field-verified reference areas for freshwater, and natural variations in

[2] Reliability: correct predictions/total stations.

bioavailability that are greater in freshwater than in marine waters for some analytes (e.g., metals).

4.3.2 Hit/No Hit Designation

A pairwise statistical comparison between the test sediment and negative control sediment will be performed following ASTM (2003) and EPA (2000) guidelines. The comparison for each test endpoint (mortality and growth) will initially be based on a statistical significance level (Type 1 error rate) of $\alpha = 0.05$. If the analysis of the toxicity test data finds that the power of the data set is low, the α level may be raised to 0.1 as suggested in ASTM guidelines (2003). The statistical analysis will be performed using Biostat 2.0 (USACE 1998) according to methods specified by EPA (2000) and ASTM (2003).

For each endpoint, the initial hit/no-hit definition will be based on a statistical difference alone. Further refinement of the hit/no-hit definition will be determined in future meetings between LWG and EPA. One alternative hit/no-hit designation that could be used is based on Ecology's (2002) proposed freshwater Sediment Quality Standards (SQS; the level above which minor adverse effects are expected) or clean-up screening levels (CSL; the level above which moderate or severe adverse effects are expected). The minimum detectable difference (MDD) for each endpoint at the SQS and CSL levels will be based on the values proposed by Ecology (2002) (Table 4-1).

Table 4-1. Proposed SQS and CSL levels based on MDD values

Test and Endpoint	SQS	CSL
Hyalella azteca 28-day mortality	T – C > 10%	T – C > 25%
Hyalella azteca 28-day growth	T/C < 0.75	T/C < 0.6
Chironomus tentans 10-day mortality	T – C > 10%	T – C > 25%
Chironomus tentans 10-day growth	T/C < 0.8	T/C < 0.7

Source: Ecology (2002)
T – test sediment value
C – negative control sediment value
Note: To be considered a test failure at either the SQS or CSL level, the test response must also be statistically significantly different from the negative control response.

4.3.3 Upstream Ambient Stations

In addition to the negative controls, an agreed-upon number of ambient sediment samples will be collected upstream of the ISA. The number and location of upstream ambient samples will be determined in consultation with EPA. Both sediment chemical analyses and toxicity tests (28-day *H. azteca* and 10-day *C. tentans*) will be conducted on samples taken from the upstream locations. This information will be used in the risk characterization step in the ERA to place the results from the ISA in regional context.

5.0 QA/QC PROCESS FOR DATA

This section presents a summary of the quality assurance/quality control (QA/QC) procedures to be used in collecting samples and in conducting sediment chemical analyses and toxicity tests. Full details of the QA/QC procedures proposed for use in this study are presented in the Round 2 Quality Assurance Project Plan (QAPP) (Integral and Windward 2004b).

5.1 SEDIMENT CHEMISTRY

Quality control samples will be prepared in the field and at the laboratory to monitor the bias and precision of the sample collection and chemical analysis procedures. Field QC samples for this study will include field replicates, field splits, equipment rinsate blanks, trip blanks, and temperature blanks. Laboratory QC samples will include matrix spike samples, matrix spike duplicates or laboratory duplicates, and method blanks. The frequency of analysis for laboratory control samples will be one for every 20 samples or one per extraction batch, whichever is more frequent. Extensive and detailed requirements for laboratory QC procedures, including calibration frequency, control limits, and requirements for corrective actions, are provided in the method protocols. As required for EPA SW-846 methods, performance-based control limits have been established by the laboratory. These and all other control limits specified in the method descriptions will be used by the laboratory to establish the acceptability of the data or the need for reanalysis of the samples. For further details on the QC procedures for chemical analyses see the Round 2 QAPP (Integral and Windward 2004b).

Field data will be verified during preparation of the sediment samples and chain-of-custody forms. Field data and chain-of-custody forms will be reviewed by the field coordinator after the field effort is complete. After field data are entered into the project database, 100% verification of the entries will be completed to ensure the accuracy and completeness of the database.

Chemistry data verification involves verifying that correct procedures were followed and that calculations were completed correctly and checking the transcriptions of the laboratory data. Data validation involves evaluating the quality and usability of the data in the context of project objectives. The first data package generated for each analytical method will be fully validated. If no problems are encountered during the validation of the data package, full validation will be completed at a rate of approximately 10% of the samples analyzed by each method. Data verification and validation will be completed by an independent validation firm.

5.2 TOXICITY TESTS

5.2.1 Verification of Qualifications for Selected Testing Laboratory

Identification of the overall process for testing sediment toxicity included three main QA/QC steps: (1) identification of the appropriate test through consensus among EPA Region 10 and their partners, outside experts from the USGS and the USACE, and the LWG; (2) identification of a contract laboratory agreed upon by EPA, LWG, and the USGS and USACE experts; and (3) a comparison of tests to be run by the LWG contract laboratory and the Columbia Environmental Research Center (CERC) in Columbia, Missouri. The objectives of this evaluation is to confirm that the sediment used by the contract laboratory is appropriate for use as the negative control and that there are no performance issues associated with conducting the toxicity tests in low hardness water (approximately 30 mg/L hardness).

The issue of having the two laboratories perform toxicity testing on split samples is currently being discussed by EPA and LWG. EPA will provide language on this issue and the text will be incorporated into this document after consensus has been reached.

In addition, an independent third party will review all of the test protocols prior to commencement of the sediment toxicity study to ensure that all laboratory test protocols are up to date and include any recently published modifications. Any discrepancies or issues in connection with the protocols will be resolved before testing is initiated. The independent party will also conduct laboratory audits during sediment testing. Any problems or deviations from the established protocols will be identified and addressed.

5.2.2 Laboratory Quality Assurance

Both sediment toxicity tests will incorporate standard QA/QC procedures for evaluation of the validity of the test results. Standard QA/QC procedures include the use of negative and positive controls and the periodic measurement of water quality during testing. The laboratory technicians performing the tests are responsible for ensuring that the appropriate procedures have been followed during the testing. The project QA/QC coordinator is responsible for ensuring that all testing performed by the laboratory meets the test acceptability criteria, is properly documented, and complete, and satisfies the project data quality objectives. The laboratory will perform the first data reduction by calculating average survival and biomass for each test sediment and the negative controls. An internal review of the data will be performed by the laboratory's QA/QC officer. A 100% external review and validation process will be performed on the toxicity test data by an independent validation firm. The electronic toxicity test data will be formatted and checked for QA/QC prior to being imported into the database. The database uses a series of stored routines to verify the integrity of the toxicity test data. For further details on the QC procedures for sediment toxicity testing, refer to the Round 2 QAPP (Integral and Windward 2004b).

6.0 DATA MANAGEMENT

This section defines the purpose of data management and describes how data are controlled from collection through use to final archiving.

6.1 OVERVIEW OF DATA MANAGEMENT PROCESS AND SYSTEM

Toxicity, analytical, and physical data will be managed in a relational database that will function in parallel with LWG database managed by Integral. The database will also have stored procedures for exporting toxicity data to both the National Oceanic and Atmospheric Administration's (NOAA's) Query Manager database and Ecology's SedQual database. Because development of the predictive relationship approach will be an iterative process, raw toxicity data will be stored in the database, allowing for the dynamic evaluation of test responses, as opposed to just the storing of laboratory-designated hit/no-hit information.

6.2 SEDIMENT ANALYTICAL AND PHYSICAL DATA

Analytical reports from the laboratory will include QC results and any other analytical information reviewers might need to assess the data quality. Initial data reduction, evaluation, and reporting performed at the laboratory will be consistent with the Round 1 QAPP (SEA 2002). Data will be delivered in both hard-copy and electronic format to the QA manager, who will be responsible for distributing them to the data validator and submitting them for permanent archiving. Hard-copy deliverables will be similar in format and content to those required by EPA's Contract Laboratory Protocol. Electronic data deliverables must be compatible with the EQuIS data management system. Hard-copy data deliverables and documentation for all laboratory results and procedures will be archived and made available to EPA upon request.

6.3 TOXICITY TEST DATA

Toxicity test reports from the laboratory will include QC results and all raw data and any other information reviewers might need to assess the data quality. The initial calculation of toxicity endpoints and a summary of QA/QC results will be reported by the laboratory. Data will be delivered in both hard-copy and electronic format to the QA Manager, who will be responsible for distributing them to the data validator and submitting them for permanent archiving. Electronic data deliverables will include all necessary QC information and data for the calculation of endpoints. Hard-copy data deliverables and documentation for all laboratory results and procedures will be archived and made available to EPA upon request.

7.0 DEVELOPMENT OF A PREDICTIVE RELATIONSHIP

This section describes the methods that will be used to develop a predictive relationship between sediment chemistry and sediment toxicity responses. By necessity, the approach will be exploratory in nature, in that a number of different methods will be used to evaluate whether a relationship exists between measured sediment chemical concentrations and toxicity responses, and in an attempt to improve the reliability of the selected approach. The sections below describe the different approaches proposed for use in the exploratory evaluation, methods for evaluating the reliability of these approaches, and the comparative approach that will be used to select and propose a model for use in Portland Harbor.

7.1 METHODS FOR PREDICTING TOXICITY

Numerous approaches have been developed to evaluate the relationship between sediment chemistry concentrations and toxicity to benthic invertebrates and to derive sediment quality values (SQVs) or criteria as a regulatory decision-making tool. The overall objective of the predictive approach is to identify chemical SQVs that adequately predict toxicity and can be used to screen sediment data based on chemical concentrations alone. Five previously established sets of freshwater SQVs will be evaluated from among those that are commonly used nation-wide or are emerging as useful tools in this region, including: 1) SQVs derived using database percentiles, 2) SQVs derived using consensus-based values, 3) a quotient method, 4) the floating percentile method, and 5) logistic regression analysis. A brief description of each of the selected methods is provided below.

7.1.1 Threshold Effects Levels and Probable Effects Levels

Threshold effects levels (TELs) and probable effects levels (PELs) are derived using the database percentile method. TELs are intended to represent chemical concentrations below which biological effects rarely occur. PELs are intended to represent chemical concentrations above which adverse biological effects frequently occur. Two freshwater versions of these levels were calculated by the US Geological Survey (USGS) and Environment Canada (Smith et al. 1996; CCME 1995), and the Canadian guidelines were updated in 2001 (CCME 2001). USGS used a smaller data set that focused on the Great Lakes Areas of Concern, whereas Environment Canada used a much larger database that included data from throughout North America. The latter set of TELs/PELs has the most widespread use.

TELs/PELs were derived by classifying sediment samples within each data set as either toxic or non-toxic. Chemical concentrations were considered to be associated with toxicity if the mean concentration at sites where adverse effects were observed was two or more times the mean concentration at sites where no effects were observed. If the mean chemical concentration at sites where adverse effects were observed was less than twice the mean chemical concentration at sites where no effects were observed, these data were reassigned to the no-effects distribution. The resulting distributions were placed in ascending order of concentration. TELs were calculated as the geometric mean of the 15th percentile of the effects distribution and the 50th percentile of the no-effects distribution. PELs were calculated as the geometric mean of the 50th percentile of the effects distribution and the 85th percentile of the no-effects distribution. If there were fewer than 20 data points in either distribution, a TEL/PEL was not calculated. TEL/PEL values have been developed for 8 metals, 12 individual PAHs, total PCBs, and 7 chlorinated pesticides (CCME 2001).

7.1.2 Threshold Effects Concentrations and Probable Effects Concentrations

Consensus-based SQVs have been proposed by a group of private and agency sediment researchers in an attempt to unify the wide variety of SQVs available in the literature (MacDonald et al. 2000; GLNPO 2000). Threshold effects concentrations (TECs) were derived using a group of existing freshwater SQV sets that represented levels below which adverse effects were seldom observed. TECs are considered conservative screening tools and not intended for use as clean-up goals. Similarly, probable effects concentrations (PECs) were derived using a group of existing freshwater SQV sets that represented levels above which adverse effects would be expected. If three or more published values with a similar narrative intent were available for a chemical or group of chemicals, the TEC or PEC was calculated as the geometric mean of these values. TECs and PECs have been developed for 8 metals, 10 individual PAHs, total PAHs, total PCBs, and 9 chlorinated pesticides (MacDonald et al. 2000).

7.1.3 Sediment Chemistry Quotients

Sediment chemistry quotients were developed as an approach to increase the predictive ability of the SQVs described above (Long et al. 1998). The quotient approach was originally developed for use with the PECs, as described in Section 7.1.2, but could be applied to any set of SQVs. A PEC quotient was calculated for each chemical in each sediment sample by dividing the concentration of a chemical by the PEC for that chemical. A mean quotient was then calculated for each sediment sample by adding the individual quotients for each chemical and dividing this sum by the number of chemicals. This approach weighted each of the individual chemicals and chemical classes equally. The mean quotient described by Long et al. (1988) was calculated using PECs for arsenic, cadmium, chromium, copper, lead, nickel, zinc, total PAHs, total PCBs, and total DDTs.

A second approach to calculating mean PEC quotients weighted the contribution of metals (as a group), total PAHs, and total PCBs equally (Ingersoll et al. 2001). The average PEC quotient was calculated for the seven metals listed above. A mean quotient was then calculated for each sediment sample by adding the average quotient for metals, the quotient for total PAHs, and the quotient for total PCBs, and then dividing this sum by three. The quotient for total DDTs is not used in this calculation.

A modification of the quotient approach using quotient sums rather than mean quotients will be evaluated as part of the exploratory approach to determine whether it improves the reliability of the quotients. It is important to note that the quotient derived by any of these methods is not equivalent to a hazard index. The specific quotient value that would be set as an SQV for sediment toxicity would be based on its ability to reliably predict the presence or absence of toxicity in Portland Harbor, and may be different for different data sets.

7.1.4 Floating Percentile Method (Washington State Freshwater Standards)

Freshwater apparent effects thresholds (AETs) were published in 1997 by Ecology (Ecology 1997). AETs were calculated from a paired chemistry and toxicity data set and are calculated separately for each biological test, primarily the *H. azteca* 10-day acute test and Microtox®, with various other tests for which fewer data were available. The first step was to assign the samples to either a hit or no-hit distribution based on a pairwise statistical comparison of the toxicity test results with those for associated reference samples. The samples in the hit and no-hit distributions were then arranged in ascending order of chemical concentrations for each individual chemical. Outliers were removed from the no-hit distribution, and the highest no-hit concentration for each chemical was selected as the AET for that chemical. Above this threshold, all concentrations of that chemical were associated with adverse effects. However, Ecology's 1997 AETs were found to have low reliability as predictors of adverse effects for freshwater sediments, as did the updated freshwater AETs calculated in 2002 (Ecology 2002).

The floating percentile method was developed in an effort to improve the reliability of freshwater SQVs for Washington State (Ecology 2002, 2003). The new method did not require the SQVs for all chemicals to be based on the same percentile of the hit or no-hit distribution. An optimal percentile of the data set that provides a low false negative rate[3] is selected, and then each individual chemical concentration is adjusted upward until the false positive rate[4] has decreased to its lowest possible level while retaining the same false negative rate. Using this method, most chemicals will be at or near their actual toxic range instead of a level arbitrarily assigned by a fixed percentile. It is possible to minimize both false positive and false negative errors at the same time, compared to other methods, because the method is primarily eliminating mathematical sources of error associated with the use of fixed percentiles to set SQVs for all chemicals.

SQVs were calculated using the floating percentile method for 11 metals, 16 individual PAHs, low-molecular-weight PAHs, high-molecular-weight PAHs, 4 phthalates, dibenzofuran, and total PCBs. These SQVs were derived using a large data set, primarily from western Washington and Oregon, including all of the Portland Harbor

[3] False negative rates: number of samples incorrectly predicted as no-hits/total number of hits.
[4] False positive rates: number of samples incorrectly predicted as hits/total number of no-hits.

data that existed at that time, and are currently applicable to freshwater sediments in Washington State (Ecology 2002).

7.1.5 Logistic Regression Model

The logistic regression approach was first proposed in 1999 as an alternative to threshold methods used for developing SQVs (Field et al. 1999, 2002). A large national data set consisting of over 3,000 marine/estuarine sediment samples with matched chemistry and toxicity tests results (two species of marine/estuarine amphipods) was assembled. The data were screened into three categories for each selected analyte: 1) non-toxic samples, 2) toxic samples with a chemical concentration greater than the mean concentration in the non-toxic samples, and 3) toxic samples with a chemical concentration lower than the mean concentration in the non-toxic samples. The designation as toxic was based on a statistically significant difference from the negative control and survival less than 90% (i.e., the minimum acceptable control survival). Only data from the first two categories were used in the logistic regression model. Individual logistic regression models were calculated for each chemical. The analyses were conducted using the Statistical Analyses System (SAS®) Institute's logistic procedure. The logistic model was then inverted to estimate the concentrations at which 20, 50, or 80% of the sediment samples would be predicted to be toxic for the selected chemical. The logistic regression model can also be used to combine multiple chemicals into a single curve, similar to the quotient method described in Section 7.1.3.

7.2 COMPARATIVE APPROACHES

Three approaches will be used to evaluate the methods described above. First, the reliability of existing SQVs will be compared using the Portland Harbor data set, including both existing data and the newly collected data. This evaluation will be performed for each of the existing SQV sets (i.e., TELs, PELs, PECs, PEC quotients, and the Washington State freshwater standards). In each case, all detected analytes that are included in the SQV set will be included in the reliability analysis. The analytes that are included vary among the SQV sets, and this is an important aspect of the evaluation of their potential reliability when applied to Portland Harbor. For each set of SQVs, the predicted hits/no-hits will be compared to the biological hits/no-hits at each station. The reliability of the SQV predictions will be assessed using the following reliability parameters: false negatives, false positives, sensitivity,[5] specificity,[6] efficiency,[7] and reliability. False negatives and false positives are the primary measures of predictive errors, with the other parameters related to these two. The reliability of these existing SQV sets will be compared against each other to determine which method or methods are best able to predict toxicity in Portland Harbor.

[5] Sensitivity: number of samples correctly predicted as hits/total number of hits.
[6] Specificity: number of samples correctly predicted as no-hits/total number of no-hits.
[7] Efficiency: number of samples correctly predicted as hits/total predicted hits.

The second approach will be to build two site-specific predictive models using the floating percentile calculation method and the logistic regression model. These calculations will be carried out as described in Sections 7.1.4 and 7.1.5, except that only the Portland Harbor data set will be used. The number of analytes that can effectively be included in these models will be largely determined by the amount of data for each detected analyte and the presence of an apparent relationship with toxicity, as determined by comparing the hit and no-hit distributions, as well as other screening steps (see original citations for details). These models will be compared against the results of the existing SQV sets to determine if there is an improvement in reliability.

A third approach is a mix of the two previous methods: Portland Harbor data are added to the existing regional and or national data sets, and the predictive relationship is recalculated. This approach will be used if the methods above show poor predictive reliability. National guidelines (TELs, PELs, and PECs) will not be recalculated because it is unlikely that the addition of the Portland Harbor data would make a significant difference. Therefore, this approach would involve adding the Portland Harbor data to other regional data from similar environments to enhance the data set for the floating percentile method and/or the logistic regression model if the Portland Harbor data set alone does not perform well. Reasons for this could include an incomplete range of chemical concentrations (e.g., not enough high-concentration samples) to calculate a reliable SQV, other large gaps in the data distribution, or widely varying bioavailability of key chemicals as a result of atypical matrices.

7.3 EVALUATING THE PREDICTIVE RELATIONSHIPS

The next step will be to determine which of the SQV sets or predictive models will be proposed for use in Portland Harbor. Each of the existing SQV sets, along with the site-specific SQVs and models, will be evaluated side by side using the same reliability parameters described above. For comparison purposes, a point on the logistic regression curve that is similar to the definition of adverse effects used in the other methods (e.g., no more than 20% adverse effects) will be selected, and an associated SQV set will be derived based on that point.

As a first cut, approaches will be eliminated if they exhibit poor performance relative to other similar approaches. For example, if Method A has the same false positive rate but lower false negatives than Method B, Method B will be eliminated in favor of Method A. Methods will also be eliminated if they have error rates that are clearly unacceptable to either the agencies or LWG (e.g., 50% false negatives or 70% false positives). In cases such as this, additional toxicity testing would be preferable to reliance on predictive approaches that perform poorly. Specific criteria for the acceptability of false negative and false positive rates will not be set in advance because it is difficult to determine in advance how much natural variability there will be in each data set. A level of error that may be acceptable (or realistic) in a data set with high inherent variability may be much greater than the level of error that would be desirable for a more homogeneous data set.

Based on experience performing a similar assessment for the derivation of the Washington State freshwater standards, it is likely that only one or two approaches will remain at this point, inasmuch as most of the existing SQVs do not perform well in predicting toxicity observed in historical data sets in Portland Harbor. If two or three methods that perform similarly and have acceptably low error rates remain, overall reliability will be the primary criterion used to select among them. Other factors to consider in making the final recommendation may include the use of regional or site-specific data, the scientific defensibility of the approach, and regulatory consistency with other SQVs being used in the region.

Each of these criteria is important, and additional discussion of how they might be used is provided below.

Overall Reliability. The more reliable the SQV set, the more likely the correct decisions will be made regarding toxicity at individual stations, which in turn will affect the accuracy of overall cleanup boundaries and areas selected for the site.

Use of Site-Specific Data. Approaches that use site-specific or regional data are preferred over those that use national data, assuming they have similar overall reliability. It is possible to have the same overall reliability yet different numeric criteria for specific chemicals. Approaches that use regional data are more likely to reflect the bioavailability and mixture-related effects associated with regional biogeochemistry and site sources.

Scientific Defensibility. In general, LWG wishes to use approaches that are scientifically defensible and based on the best available science. Certain approaches, such as the consensus-based approach, are not based on specific models or theories but are rather combinations of other approaches. Some of the methods described above have been extensively used and validated by field and laboratory experiments, while others are newly emerging and appear promising but have been less widely used. Other factors being equal, scientific defensibility is an important consideration and will increase the credibility of the resulting cleanup values among all of the stakeholders involved in the process.

Regulatory Consistency. The Portland Harbor Superfund Site is an area that is encompassed by a number of other regional initiatives, including dredging projects and the development of an Oregon Department of Environmental Quality (DEQ) sediment policy to unify approaches to dredging and cleanup along the Columbia River and its tributaries. It is desirable that the cleanup values selected for this site be consistent with those being used or emerging as part of the surrounding regulatory efforts.

The results of the reliability analysis for each of the existing SQV sets, as well any refinements or site-specific approaches developed, will be documented in a technical memorandum prior to completion of the RI/FS report. This technical memorandum will be distributed to LWG and the agencies for discussion and final selection of the benthic approach prior to completing the risk assessment as part of the RI/FS.

8.0 APPLICATION OF THE SELECTED APPROACH

The first step in applying the predictive approach is to determine whether it can be applied to the entire site, or whether there are areas within the ISA where it may not be sufficiently predictive. This will be determined by mapping the stations at which prediction errors (i.e., false positives and false negatives) are observed in the paired toxicity test/chemistry results. If the stations at which errors occur are evenly or randomly distributed, then the approach will be considered acceptably predictive across the entire site. If, however, the errors cluster in the vicinity of certain sites or sources, this will be an indication that in these areas, biological testing is needed to accurately assess the risks to the benthic community. The floating percentile method allows for the evaluation of errors related to individual chemicals or chemical classes. This may also be of value in identifying geographic areas for which the predictive approach may not perform well.

For areas where the predictive approach is applied, surface sediment chemistry will be compared to the selected SQVs to define areas of varying sediment quality relative to benthic organism health. This information will then be used in the risk characterization portion of the baseline risk assessment to estimate risks to benthic populations and the community as a whole. More than one method of evaluation may be used.

First, the derived SQVs may be used to identify impacted stations based on chemistry alone, where toxicity tests have not been conducted. At stations where toxicity tests have been conducted, toxicity test results will be used instead, because these are presumed to reflect actual toxicity at the station. This provides a hit or no-hit determination of adverse effect for each station, which can be used to develop cleanup boundaries.

In addition, some of the methods described above (e.g., the logistic regression model, quotient methods, floating percentile method) can be used in a more quantitative evaluation of predicted adverse effects, along a range from “not likely to be impacted” to “highly likely to be impacted” with several gradations in between. In this case, the SQVs may be used to map areas with varying degrees of predicted toxicity. This information can be added to other risk-based information, such as bioaccumulative risks, to determine the need for cleanup in each area and/or the type of cleanup that would be appropriate.

9.0 APPLICATION OF ALTERNATIVE APPROACHES

If a predictive approach cannot be established to sufficiently estimate sediment toxicity for benthic invertebrates based on sediment chemistry in either selected areas or across the entire site, an alternative approach will be necessary. Criteria for determining the predictive reliability and acceptability of the benthic approaches are discussed in Section 7.3. As discussed in Section 8, there may be specific areas within the ISA for which the selected method is not predictive, but it is also possible that no predictive method can be established for any area of the ISA. Alternative approaches to be used in either circumstance are discussed below.

9.1 ALTERNATIVE APPROACH FOR ALL PORTLAND HARBOR SEDIMENTS

Model failure could result from sampling a range of chemical concentrations that is insufficient to develop a predictive model. Toxicity testing stations were placed across a range of sediment chemistry conditions to capture the full range of sediment chemistry found in Portland Harbor, and previous investigations have indicated that sediment concentrations for various chemicals may range over as many as six orders of magnitude (PTI 1992). Therefore, the targeted and grid sediment sampling scheme outlined in the Round 2 FSP (Integral and Windward 2004a) should provide a sufficient range of chemical concentrations for the development of a predictive model. However, even given a sufficient range of chemical concentrations, there are three potential scenarios where no applicable predictive model could be established for all Portland Harbor sediments. These scenarios are discussed below.

9.1.1 Threshold Sediment Concentration

In this scenario, a predictive model cannot be established for sediment chemistry and toxicity, but toxic responses (hits) are only observed at higher chemical concentrations. There may be some threshold concentration below which sediments are not toxic; but above the threshold concentration, there may be no predictive relationship between sediment chemistry and toxicity. An example of a frequency distribution of data under this scenario is presented in Figure 9-1.

Under this scenario, another round of toxicity testing would likely be required, but it would be possible to screen out toxicity testing at sediment chemistry stations with concentrations below the threshold concentration itself or some function of the threshold as a safety factor.

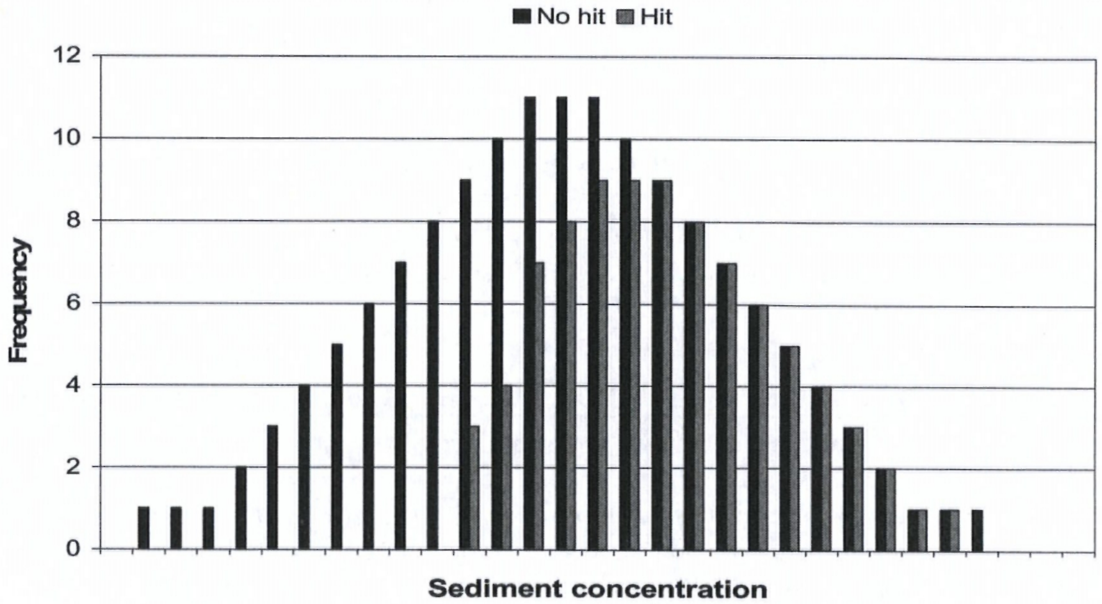

Figure 9-1. Idealized frequency distribution with toxic concentration threshold but no predictive relationship

9.1.2 No Sediment Chemistry-Toxicity Relationship

Under this scenario, a wide range of test responses is observed, but there is no apparent relationship between sediment chemistry and toxicity. If the toxicity test results fall under this scenario, a predictive model cannot be used in lieu of additional toxicity testing. Because it would not be possible to screen out any sediment chemistry stations based on chemistry alone, an agreed-upon (between EPA and LWG) number of remaining sediment stations would have to be resampled and tested for toxicity.

9.1.3 Limited Range of Toxicity Response

Finally, it can be difficult to establish a predictive approach if either no-hit or hit distributions have too few data. Because the sediment chemistry sampling includes targeted sampling around known sources and hot spots and grid sampling in other areas, the full range of sediment chemistry should be tested for toxicity. Therefore, few samples in the hit distribution and no apparent relationship between % response and sediment chemistry would suggest that despite high chemical concentrations, the sediment in Portland Harbor is generally not very toxic to benthic invertebrates (chironomids and amphipods), and what toxicity does occur may not be related to the chemicals evaluated. The two toxicity tests used are relatively sensitive, so different tests would not be expected to provide different results. However, given the expected chemical concentrations near some known sources, and previous findings of toxicity for some of those sediments, this is an unlikely scenario.

Conversely, too few data in the no-hit distribution would suggest that nearly all sediments in Portland Harbor are toxic to chironomids and amphipods. Less-sensitive toxicity tests could possibly indicate lower toxicity, but previous toxicity studies from Portland Harbor with numerous non-toxic sediment samples indicate that this scenario is also unlikely. Either of these scenarios could suggest systematic problems with the toxicity testing procedures.

There could, however, still be a relationship between toxicity and sediment chemistry if the few samples in the hit distribution are only for sediments with the highest chemical concentrations, or if the few samples in the no-hit distribution are only for sediments with the lowest chemical concentrations. In either of these cases, a predictive model will be established as discussed in Section 7. It is likely that in this case, additional regional data would be added to supplement the data set and fill in the concentration distribution gaps, as described in Section 7.2.

9.2 ALTERNATIVE APPROACH FOR SUBSET OF PORTLAND HARBOR SEDIMENTS

As discussed in Section 8, a predictive approach may be developed for most of the sediment samples within Portland Harbor, but there could be areas for which a predictive approach would not apply. This could occur for several possible reasons:

- The physical form of a contaminant (e.g., TBT in paint chips) might limit its bioavailability in one area compared to the rest of the ISA.
- Physical conditions that affect toxicity (e.g., TOC or sediment grain size) could vary significantly from one area in the ISA to another.
- The localized presence of a single contaminant could drive toxicity in several isolated samples but not throughout the ISA (e.g., DDT).

Within the areas of the ISA for which the selected approach is not predictive, two possible alternative approaches could be used. First, depending on an analysis of the chemical distributions, one of the methods identified in Section 9.1 could be used, just as it would be for the entire site.

Alternatively, if the area in which the approach is not predictive has enough data and a specific geographic or geochemical pattern that is likely associated with the errors observed can be identified, the area could be divided into subareas and the predictive approach run separately for each subarea. For example, if physical conditions and associated bioavailability appeared very different within the navigation channel compared to areas along the banks, the site could be stratified and separate models or SQVs developed for each subarea. As another example, if the existing standard quotient method (which doesn't incorporate DDTs) was used for the majority of the site, an alternative quotient method that takes DDTs into account in areas where they are elevated could be developed.

10.0 REFERENCES

ASTM. 2001. Standard test methods for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. ASTM standard method no. E 1706-00. American Society for Testing and Materials, Philadelphia, PA.

ASTM. 2003. Standard test methods for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. ASTM standard method no. E1706-00. Second edition. American Society for Testing and Materials, Philadelphia, PA.

CCME. 1995. Protocol for the Derivation of Canadian Sediment Quality Guidelines for the Protection of Aquatic Life. CCME EPC-98E. Canadian Council of Ministers of the Environment, Winnipeg, Canada.

CCME. 2001. Canadian Sediment Quality Guidelines for the Protection of Aquatic Life. Updated Summary Tables. Canadian Council of Ministers of the Environment, Winnipeg, Canada.

Ecology. 1997. Creation and analysis of freshwater sediment quality values in Washington State. Washington Department of Ecology, Environmental Investigation and Laboratory Services Program, Olympia, WA.

Ecology. 2002. Development of freshwater sediment quality values for use in Washington State. Phase I task 6: final report. Publication no. 02-09-050. Toxics Cleanup Program, Washington Department of Ecology, Olympia, WA.

Ecology. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: development and recommendation of SQVs for freshwater sediments in Washington State. Publication no. 03-09-088. Toxics Cleanup Program, Washington Department of Ecology, Olympia, WA.

EPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. Second edition. EPA 600/R-99/064.

Field J, Norton S, MacDonald D, Severn C, Ingersoll C. 1999. Beyond thresholds: using logistic regression models to estimate the probability of toxicity from sediment chemistry. SETAC poster.

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn C, Smorong DE, Lindskoog RA. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ Toxicol Chem 9:1993-2005.

GLNPO. 2000. Prediction of Sediment Toxicity using Consensus-Based Freshwater Sediment Quality Guidelines. Great Lakes National Program Office, Chicago, IL. EPA 905-R-00-007.

Hart Crowser. 2002. Lower Willamette River reference area study volume 1. Prepared for US Army Corps of Engineers. Portland, OR.

Ingersoll CG, MacDonald DD, Wang N, Crane JL, Field LJ, Haverland PS, Kemble NE, Lindskoog RA, Severn C, and DE Smorong. 2001. Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. Arch Environ Contam Toxicol 41:8-21.

Ingersoll, CG, D MacDonald, S Norton, WG Brumbaugh, B Johnson, N Kem, J Kunz, T May, N Wang, J Smith, D Sparks, and S Ireland. 2002. Toxicity assessment of sediments from the Grand Calumet and Indiana Harbor canal in Northwestern Indiana, USA. Arch Environ Contam Toxicol 43:156-167.

Integral, Windward. 2004a. Portland Harbor remedial investigation/feasibility study round 2 field sampling plan: sediment sampling and benthic toxicity testing. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA and Windward Environmental LLC, Seattle, WA.

Integral, Windward. 2004b. Portland Harbor remedial investigation/feasibility study round 2 project quality assurance project plan. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA and Windward Environmental LLC, Seattle, WA.

Integral, Windward, Kennedy/Jenks, Anchor, Groundwater Solutions. 2004. Portland Harbor remedial investigation/feasibility study programmatic work plan. Revised. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA; Windward Environmental LLC, Seattle, WA; Kennedy/Jenks Consultants, Federal Way, WA; Anchor Environmental, LLC, Seattle, WA; Groundwater Solutions, Inc., Portland, OR.

Kleinfelder. 2004. Gunderson, Inc. Area 2 - sandy beach area upland source evaluation. Kleinfelder, Inc., Beaverton, OR.

LCRMA. 1998. Dredged material evaluation framework guidance for the lower Columbia River Management Area. Draft. US Army Corps of Engineers, Seattle District, Portland District, Northwestern Division; US Environmental Protection Agency, Region 10; Washington State Department of Ecology; Oregon Department of Environmental Quality; Washington State Department of Natural Resources, Seattle, WA and Portland, OR.

Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environ Toxic Chem 17:714-727.

MacDonald DD, Ingersoll CB, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39:20-31.

PTI. 1992. McCormick & Baxter Creosoting Company remedial investigation report: Vol I. Prepared for Oregon Department of Environmental Quality. PTI Environmental Services, Inc., Bellevue, WA.

SEA. 2002. Portland Harbor remedial investigation/feasibility study round 1 quality assurance project plan. Prepared for Lower Willamette Group. Striplin Environmental Associates, Inc., Olympia, WA.

Smith SL, MacDonald DD, Keenleyside KA, Ingersoll CG, and Field LJ. 1996. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. J. Great Lakes Research 22(3):624-638.

USACE. 1998. Biostat 2.0 users guide. Review draft. Seattle District, US Army Corps of Engineers, Seattle, WA.

Windward, Kennedy/Jenks, Groundwater Solutions. 2003. Framework for evaluating exposure to the benthic community and humans from chemicals transported in groundwater. Draft. Prepared for the Lower Willamette Group. Windward Environmental LLC, Seattle, WA; Kennedy/Jenks Consultants, Federal Way, WA; Groundwater Solutions, Inc., Portland, OR.